WHITE PAPER

DiffServ—The Scalable End-to-End QoS Model

Executive Summary:

The Internet is changing every aspect of our lives—business, entertainment, education, and more. Businesses use the Internet and Web-related technologies to establish Intranets and Extranets that help streamline business processes and develop new business models.

Behind all this success is the underlying fabric of the Internet: the Internet Protocol (IP). IP was designed to provide best-effort service for delivery of data packets and to run across virtually any network transmission media and system platform. The increasing popularity of IP has shifted the paradigm from “IP over everything,” to “everything over IP.” In order to manage the multitude of applications such as streaming video, voice over IP, e-commerce, ERP, and others, a network requires quality of service (QoS) in addition to best-effort service. Different applications have varying needs for delay, delay variation (jitter), bandwidth, packet loss, and availability. These parameters form the basis of QoS. The IP network should be designed to provide the requisite QoS to applications.

For example, VoIP requires very low jitter, and a one-way delay in the order of 100 milliseconds, and guaranteed bandwidth in the range of 8Kbps –> 64Kbps, dependent on the codec used. On the other hand, a file transfer application, based on ftp, doesn’t suffer from jitter, while packet loss will be highly detrimental to the throughput.

To facilitate true end-to-end QoS on an IP-network, the Internet Engineering Task Force (IETF) has defined two models: Integrated Services (IntServ) and Differentiated Services (DiffServ). IntServ follows the signaled-QoS model, where the end-hosts signal their QoS need to the network, while DiffServ works on the provisioned-QoS model where network elements are set up to service multiple classes of traffic, with varying QoS requirements. Both models can be driven off a policy base, using the CoPS (Common Open Policy Server) protocol. Cisco IOS® software supports both the IntServ and DiffServ models of QoS, along with an optional CoPS-client functionality.

IntServ provides for a rich end-to-end QoS solution, by way of end-to-end signaling, state-maintenance (for each RSVP-flow and reservation), and admission control at each network element. DiffServ, on the other hand, addresses the clear need for relatively simple and coarse methods of categorizing traffic into different classes, also called class of service (CoS), and applying QoS parameters to those classes. To accomplish this, packets are first divided into classes by marking the type of service (ToS) byte in the IP header. A 6-bit bit-pattern (called the Differentiated Services Code Point [DSCP]) in the IPv4 ToS Octet or the IPv6 Traffic Class Octet is used to this end as shown in Figures 1, 2, and 3.

Figure 1 IPv4 and IPv6 Headers

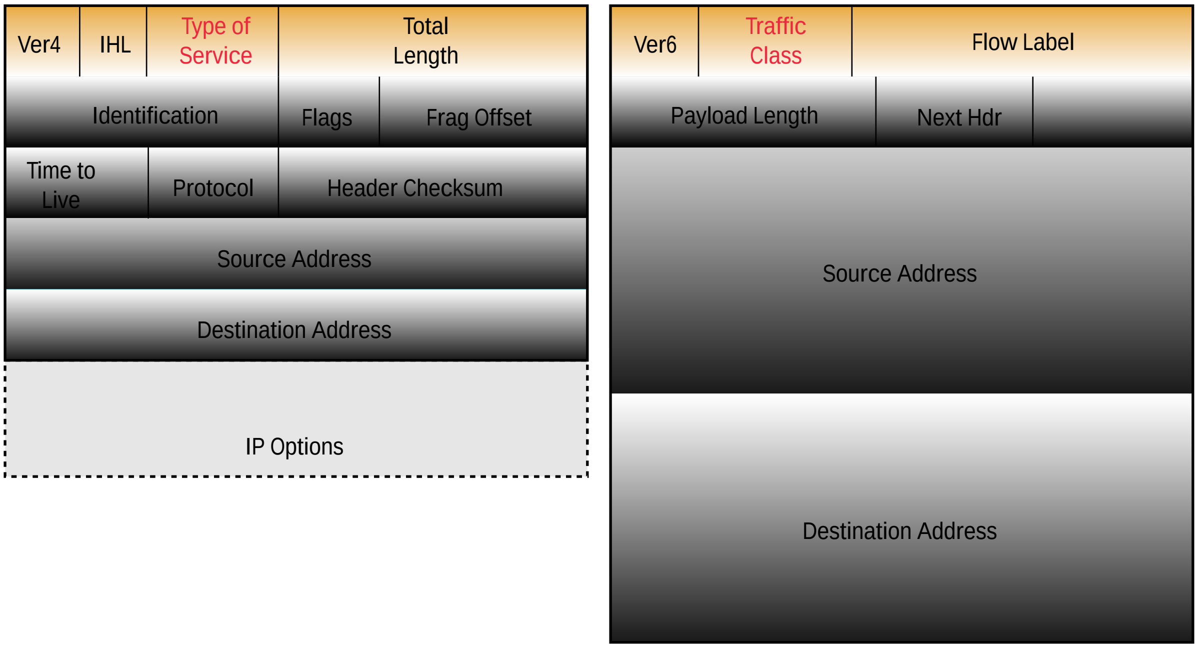
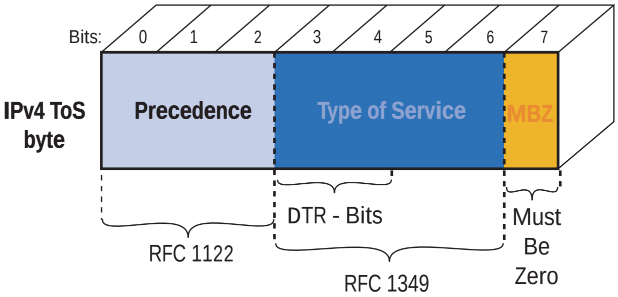


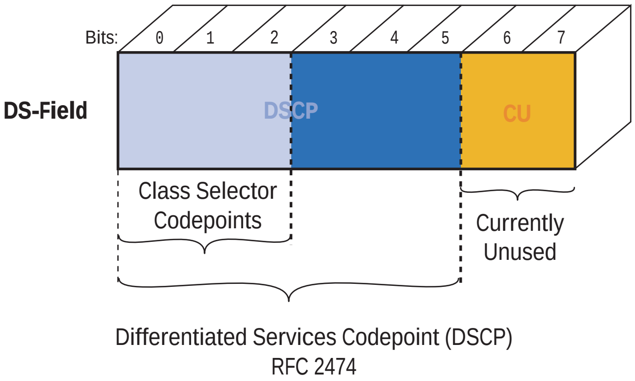
Figure 2 The Original IPv4 ToS Byte



Bits (0-2): IP-Precedence Defined
111 - Network Control
110 - Internetwork Control
101 - CRITIC/ECP
100 - Flash Override
011 - Flash
101 - Immediate
001 - Priority
000 - Routine

Bits (3-6): The Type of Service Defined
0000 (all normal)
1000 (minimize delay)
0100 (maximize throughput)
0010 (maximize reliability)
0001 (minimize monetary cost)

Figure 3 DiffServ Codepoint Field


Bits:
0
1
2
3
4
5
6
7
DS-Field
DSCP
CU
Class Selector Codepoints
Currently Unused
Differentiated Services Codepoint (DSCP)
RFC 2474

Once packets are classified at the edge of the network, specific forwarding treatments, formally called PHB (Per-Hop Behavior), are applied on each network element, providing the packet the appropriate delay-bound, jitter-bound, bandwidth, etc. This combination of packet marking and well-defined PHBs results in a scalable QoS solution for any given packet, and thus any application. Thus, in DiffServ, signaling for QoS is eliminated, and the number of states required to be kept at each network element is drastically reduced, resulting in a coarse-grained, scalable and end-to-end QoS solution.

Why Do We Need DiffServ?

IntServ, Its Strengths and Shortcomings

The IETF defined models, IntServ and DiffServ, are simply two ways of considering the fundamental problem of providing QoS for a given IP packet. The IntServ model relies on the Resource Reservation Protocol (RSVP) to signal and reserve the desired QoS for each flow in the network. A flow is defined as an individual, unidirectional data stream between two applications, and is uniquely identified by the 5-tuple (Source IP address, Source Port#, Destination IP address, Destination Port#, and the Transport Protocol). Two types of service can be requested via RSVP (assuming all network devices support RSVP along the path from the source to the destination). The first type is a very strict guaranteed service that provides for firm bounds on end-to-end delay and assured bandwidth for traffic that conforms to the reserved specifications. The second type is a controlled load service that provides for a better than best effort and low delay service under light to moderate network loads. Thus, it is possible (at least theoretically) to provide the requisite QoS for every flow in the network, provided it is signaled using RSVP, and the resources are available. However, there are several practical drawbacks to this approach:

- Every device along a packet's path, including the end systems like servers and PCs, need to be fully aware of RSVP and capable of signaling the required QoS.
- Reservations in each device along the path are "soft," which means they need to be refreshed periodically, thereby adding to the traffic on the network and increasing the chance that the reservation may time out if refresh packets are lost. Though some mechanisms alleviate this problem, it adds to the complexity of the RSVP solution.
- Maintaining soft-states in each router, combined with admission control at each hop adds to the complexity of each network node along the path, along with increased memory requirements, to support large number of reservations.
- Since state information for each reservation needs to be maintained at every router along the path, scalability with hundreds of thousands of flows through a network core becomes an issue.

On Layer2 QoS Mechanisms

Before the IETF defined IP (Layer3) QoS methods, the ITU-T (International Union for Telecommunications, Telecommunications), the Asynchronous Transfer Mode (ATM) Forum, and the Frame-Relay Forum (FRF) had already arrived at standards to do Layer2 QoS in ATM and Frame-Relay networks. The ATM standards define a very rich QoS infrastructure by supporting traffic contracts, many adjustable QoS knobs (such as Peak Cell Rate [PCR], Minimum Cell Rate [MCR], and so on), signaling, and Connection Admission Control (CAC) [Ref-A]. Frame Relay, on the other hand provides for a simpler yet rich set of mechanisms to provide for a Committed Information Rate (CIR), Congestion Notification, and the recently introduced Frame-Relay Fragmentation (FRF.12) [Ref-B].

Though these rich QoS mechanisms exist in Layer2 transport technologies, true end-to-end QoS is not achievable, unless a Layer3 solution is overlaid. Service providers offering both ATM/Frame-Relay and IP services want to provide robust QoS solutions to customers. Mapping Layer3 QoS to Layer2 QoS is the first step toward achieving a complete solution that does not depend on any specific Layer2 technology. Both IntServ and DiffServ can be implemented over QoS-aware transports such as ATM and Frame-Relay. For example, the IntServ controlled-load service can implemented using RSVP, over an ATM VBR-rt (Variable Bit Rate, Real-Time) Switched Virtual Circuit (SVC). With DiffServ, packets marked with different ToS-byte values can be sent over different ATM PVCs or SVCs. As an example, high priority traffic may go over a VBR-nrt VC, and all other traffic may go over an Available Bit Rate (ABR) VC, with the VBR VC capable of preempting the ABR VC in case of congestion or failure. Similarly, Frame-Relay Traffic Shaping (FRTS) (slowing down the rate of transmission by buffering, in response to congestion notification by the FR switches), FRF.12 (packet fragmentation and interleaving on low speed FR links), and other mechanisms can be used to complement IP QoS.

Thus, a true end-to-end QoS solution comprises both Layer3 and Layer2 QoS, and is media independent. Thus, introduction of a Gigabit Ethernet link somewhere along the packet's path poses no problem to deliver QoS, as the Layer3 QoS is still preserved, and can even be enhanced by mapping to the 802.1p (User-Priority) QoS mechanism on Ethernet (RFC-1349). Cisco IOS QoS focuses on delivering exactly this model—inter-operability/mappings between Layer2 and Layer3 QoS over IP, ATM, Frame-Relay, Packet over SONET (POS), Ethernet, etc.

A Simpler Middle Ground

Since per-flow QoS is difficult to achieve end-to-end in a network without adding significant complexity, cost, and introducing scalability issues, it naturally leads one to think about classifying flows into aggregates (classes), and providing appropriate QoS for the aggregates. For example, all TCP flows could be grouped as a single class, and bandwidth allocated for the class, rather than for the individual flows. In addition to classifying traffic, signaling and state maintenance requirements on each network node should me minimized. The IETF realized this, and defined a mechanism to use the type of service (ToS) field in the IPv4 header to prioritize packets as shown in Figures 1, 2 and 3. Once packets are marked with the appropriate priority/IP precedence bits, any network node along the packet's path knows the relative importance (priority level) of the packet, and can apply preferential forwarding to packets of packets of higher priority levels.

The ToS/IP Precedence Solution

The IPv4 ToS byte in the IP-header as shown in Figure 1 is defined in Figure 2. The three precedence bits are mainly used to classify packets at the edge of the network into one of the eight possible categories listed in Figure 2. Packets of lower precedence (lower values) can be dropped in favor of higher precedence when there is congestion on a network. Further, each packet may be marked to receive one of two levels of delay, throughput, and reliability (the DTS bits) in its forwarding (RFC-791). However, RFC-1349 redefines these three bits, and adds the 7th bit in the byte as well for designating the ToS request for the packet (Figure 2), in addition to its priority. It may appear that this simple scheme has all the ingredients necessary to support scalable IP QoS in a network. However, this scheme has a few crucial limitations/missing components:

- The IP-Precedence scheme allows only specification of relative priority of a packet. It has no provisions to specify different drop precedence for packets of a certain priority. For example, a network administrator may want to specify both HTTP and Telnet traffic as high-priority packets. However, when there is congestion he/she wants the telnet packets to be dropped, before the HTTP (a valid reason may be because HTTP are carrying e-commerce traffic, while the Telnet packets are carrying user-sessions within the company for their enterprise resource planning [ERP] application). It is not possible to do this with the IP-Precedence scheme.
- The 3 bits restrict the number of possible priority classes to eight. Further, the Network Control and Internetwork Control classes are usually reserved for router-generated packets such as routing updates, ICMP messages, etc. This is done to protect the packets that are necessary for the health of the network. However, this cuts down the usable classes for production traffic to 6.
- Neither IP-Precedence, nor the DTS bits (bits 3,4,5—the original type of service subfield) are implemented consistently by network vendors today. In addition, RFC-1349 redefines the type of service subfield, by utilizing bits 3,4,5, and 6, and eliminating the DTS concept.

All of the above reduce the chances of successfully implementing end-to-end QoS using this scheme.

The Solution

The Differentiated Services Architecture:

The IETF completed the request for comments (RFCs) for DiffServ toward the end of 1998. As stated in the DiffServ working group objectives [Ref-C], “There is a clear need for relatively simple and coarse methods of providing differentiated classes of service for Internet traffic, to support various types of applications, and specific business requirements. The differentiated service approach to providing quality of service in networks employs a small, well-defined set of building blocks from which a variety of aggregate behaviors may be built. A small bit-pattern in each packet, in the IPv4 ToS octet or the IPv6 Traffic Class octet, is used to mark a packet to receive a particular forwarding treatment, or per-hop behavior, at each network node. A common understanding about the use and interpretation of this bit-pattern is required for inter-domain use, multi-vendor interoperability, and consistent reasoning about expected aggregate behaviors in a network. Thus, the working group has standardized a common layout for a six-bit field of both octets, called the DS field. RFC 2474 and RFC 2475 define the architecture, and the general use of bits within the DS field (superseding the IPv4 ToS octet definitions of RFC 1349).”

In order to deliver end-to-end QoS, this architecture (RFC-2475) has two major components—Packet Marking using the IPv4 ToS byte, and Per Hop Behaviors (PHBs).

Packet Marking

Unlike the IP-Precedence solution, the ToS byte is completely redefined [Figure3]. Six bits are now used to classify packets. The field is now called the DS (Differentiated Services) Field, with two of the bits unused (RFC-2474). The 6 bits replace the three IP-Precedence bits, and is called the Differentiated Services Codepoint (DSCP). With DSCP, in any given node, up to 64 different aggregates/classes can be supported (2^6). All classification and QoS revolves around the DSCP in the DiffServ model.

Per Hop Behaviors (PHBs)

Now that packets can be marked using the DSCP, how do we provide meaningful classification on flows (CoS), and provide the QoS that is needed? First, the collection of packets that have the same DSCP value (also called a Codepoint) in them, and crossing in a particular direction is called a Behavior Aggregate (BA). Thus, packets from multiple applications/sources could belong to the same BA. Formally, RFC-2475 defines a PHB as the externally observable forwarding behavior applied at a DS-compliant node to a DS behavior aggregate. In more concrete terms, a PHB refers to the packet scheduling, queuing, policing, or shaping behavior of a node on any given packet belonging to a BA, and as configured by an SLA or policy. To date, four standard PHBs are available to construct a DiffServ-enabled network and achieve coarse-grained end-to-end CoS and QoS:

The Default PHB (Defined in RFC-2474)

The default PHB essentially specifies that a packet marked with a DSCP value (recommended) of ‘000000’ gets the traditional best effort service from a DS-compliant node (a network node that complies to all the core DiffServ requirements). Also, if a packet arrives at a DS-compliant node and its DSCP value is not mapped to any of the other PHBs, it will get mapped to the default PHB.

Class-Selector PHBs (Defined in RFC-2474)

To preserve backward compatibility with the IP-Precedence scheme, DSCP values of the form ‘xxx000’, where x is either 0 or 1 are defined. These Codepoints are called Class-Selector Codepoints. Note that the default Codepoint is also is also a Class-Selector Codepoint (‘000000’). The PHB associated with a Class-Selector Codepoint is a Class-Selector PHB. These PHBs retain almost the same forwarding behavior as nodes that implement IP-Precedence based classification & forwarding. As an example, packets with a DSCP value of ‘110000’ (IP-Precedence 110) have a preferential forwarding treatment (scheduling, queuing, etc.) as compared to packets with a DSCP value of ‘100000’ (IP-Precedence 100). These PHBs ensure that DS-compliant nodes can co-exist with IP-Precedence aware nodes, with the exception of the DTS bits.

Expedited Forwarding (EF) PHB (Defined in RFC-2598)

Just as RSVP, via the IntServ model, provides for a guaranteed bandwidth service, the EF PHB is the key ingredient in DiffServ for providing a low-loss, low-latency, low-jitter, and assured bandwidth service. Applications such as voice over IP (VoIP), video, and online trading programs require such a robust network-treatment. EF can be implemented using priority queuing, along with rate limiting on the class (formally, a Behavior Aggregate). Although EF PHB when implemented in a DiffServ network provides a premium service, it should be specifically targeted toward the most critical applications, since if congestion exists, it is not possible to treat all or most traffic as high priority. EF PHB is especially suitable for applications (like VoIP) that require very low packet loss, guaranteed bandwidth, low delay, and low jitter. The recommended DSCP value for EF is ‘101110’(RFC-2474).

Assured Forwarding (AFxy) PHB (Defined in RFC-2597)

The rough equivalent of the IntServ Controlled Load Service is the Assured Forwarding PHB. It defines a method by which Behavior Aggregates can be given different forwarding assurances. For example, traffic can be divided into gold, silver, and bronze classes, with gold being allocated 50 percent of the available link bandwidth, silver 30 percent, and bronze 20 percent. The AFxy PHB defines four AFx classes; namely, AF1, AF2, AF3, and AF4. Each class is assigned a certain amount of buffer space and interface bandwidth, dependent on the SLA with the Service Provider/policy. Within each AFx class, it is possible to specify 3 drop precedence values. Thus, if there is congestion in a DS-node on a specific link, and packets of a particular AFx class (say AF1) need to be dropped, packets in AFxy will be dropped such that the dP(AFx1) <= dP(AFx2) <= dp(AFx3), where dP(AFxy) is the probability that packets of the AFxy class will be dropped. Thus, the subscript “y” in AFxy denotes the drop precedence within an AFx class. In our example, packets in AF13 will get dropped before packets in AF12, before packets in AF11. This concept of drop precedence is useful, for example, to penalize flows within a BA that exceed the assigned bandwidth. Packets of these flows could be re-marked by a policer to a higher drop precedence. Table 1 shows the DSCP values for each class, and drop precedence. And AFx class can be denoted by the DSCP ‘xyzab0’, where ‘xyz’ is 001 / 010 / 011 / 100, and ‘ab’ represents the drop precedence bits (RFC-2597).

Table 1 DiffServ AF Codepoint Table

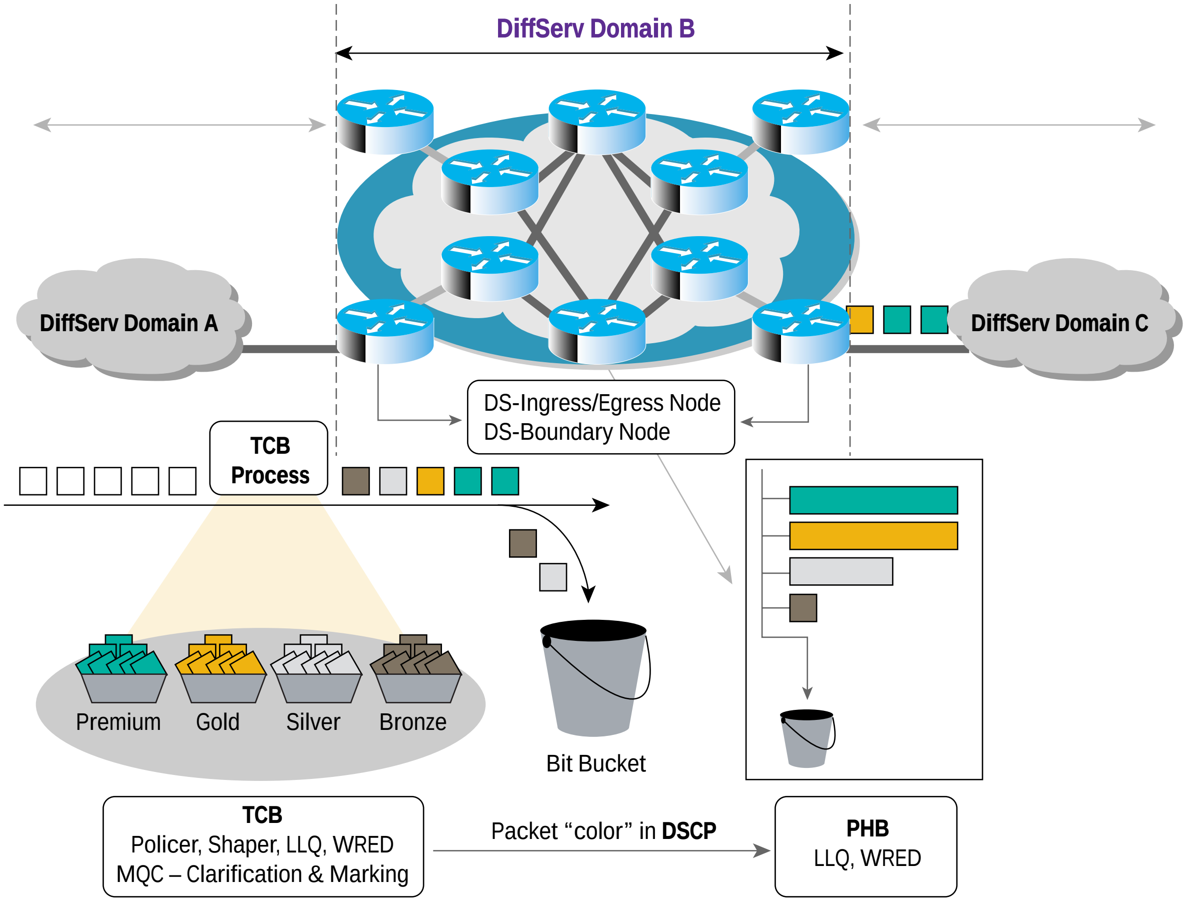
DROP Precedence	Class #1	Class #2	Class #3	Class #4
Low Drop Prec	(AF11) 001010	(AF21) 010010	(AF31) 011010	(AF41) 100010
Medium Drop Prec	(AF12) 001100	(AF22) 010100	(AF32) 011100	(AF42) 100100
High Drop Prec	(AF13) 001110	(AF23) 010110	(AF33) 011110	(AF43) 100110

Baking the DiffServ Pie

Baking the perfect pie requires both the best ingredients, as well as a great recipe. The DiffServ Pie, (the DS-Region) is composed of one or more DS-Domains, possibly under multiple administrative authorities. Each DS-Domain in turn is prepared by using the DSCP and the different PHBs. The secret to the whole recipe is the Service Level Agreement (SLA), or policy.

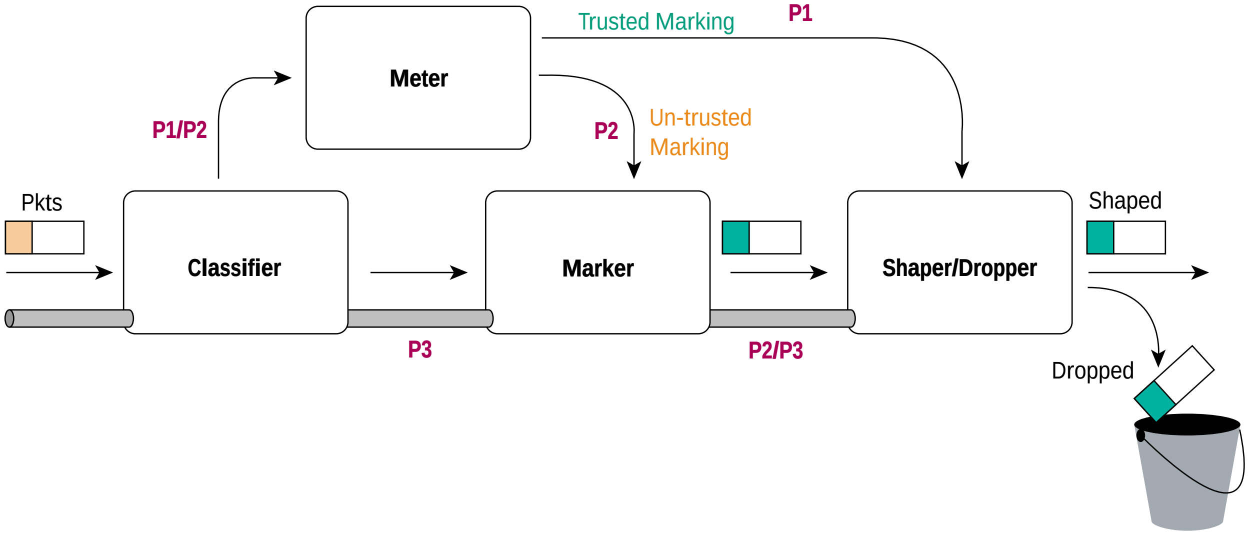
Figure 4 gives a pictorial overview of this end-to-end architecture. For true QoS, the entire IP path that a packet travels must be DiffServ enabled. An example service policy might be that EF gets 10 percent, Gold 40 percent, Silver 30 percent, Bronze 10 percent, and Best Effort traffic (default class/PHB) the remaining 10 percent of the bandwidth. Gold, Silver, and Bronze could be mapped to AF classes AF1, AF2, and AF3 for example. This can be enforced in any part of the cloud, including end-to-end.

Figure 4 DiffServ Architecture



A DS-Domain itself is made up of DS Ingress nodes, DS Interior nodes (in the core), and DS Egress nodes. Further, an Ingress or Egress node might be a DS Boundary Node, connecting two DS Domains together. Functionally all DS Ingress and Egress nodes can be categorized as a Boundary node, since they act as a demarcation point between the DS-Domain and the non-DS-aware (L2-LAN, etc.) network. Typically, the DS Boundary node performs traffic conditioning. A traffic conditioner typically classifies the incoming packets into pre-defined aggregates, meters them to determine compliance to traffic parameters (and determines if the packet is in profile, or out of profile), marks them appropriately by writing/re-writing the DSCP, and finally shapes (buffers to achieve a target flow rate) or drops the packet in case of congestion. Figure 5 illustrates the typical traffic conditioner at the edge of a DS-Domain. A DS Internal node enforces the appropriate PHB by employing policing or shaping techniques, and sometimes re-marking out of profile packets, depending on the policy or the SLA.

Figure 5 DiffServ Traffic Conditioner Block (TCB)



Classifier: selects a packet in a traffic stream based on the content of some portion of the packet header.
Meter: checks compliance to traffic parameters (e.g., Token Bucket) and passes results to marker and shaper/dropper to trigger action for in/out-of-profile packets.
Marker: writes/rewrites the DSCP value
Shaper: delay some packets for them to be compliant with the profile.

DiffServ in Cisco IOS software Today

The Mechanisms

Today, the DiffServ model only defines the use of the DSCP and the four PHBs. The PHBs simply describe the forwarding behavior of a DS-compliant node. The model does not specify how these PHBs may be implemented. A variety of queuing, policing, metering, and shaping techniques may be used to affect the desired traffic conditioning and PHB.

The Modular QoS CLI (MQC)

The MQC is a provisioning mechanism in Cisco IOS software, which allows for separation of packet classification (class-maps), from policies (policy-maps) applied on the defined classes, from the application of those policies on interfaces and sub-interfaces (service-policy) [Ref-D]. The MQC forms the basis for provisioning DiffServ, and all the QoS mechanisms are part and parcel of the class-maps (classification), or policy-maps (policing, shaping, queuing, congestion avoidance, packet marking, Layer2 CoS marking, etc.).

Packets entering a DiffServ Domain (DS-Domain) can be metered, marked, shaped, or policed to implement traffic policies (as defined by the administrative authority). In Cisco IOS software, classifying, and marking is done using the MQC's class-maps. Metering is done using a token bucket algorithm, shaping is done using Generic Traffic Shaping (GTS) or Frame Relay Traffic Shaping (FRTS), and policing is done using class-based Policing/Committed Access Rate (CAR). On the value add side, Cisco also provides for the Per-Class Accounting MIB, wherein statistics for each class (regardless of congestion) can be gleaned for management purposes. See Table 2 for the list of features in various categories.

Table 2 DiffServ Mechanisms in Cisco IOS software

Classifier	Conditioner	Forwarding	PHB	Accounting
In MQC*, classify based on (a) Interface (b) MAC address (c) ACL (d) NBAR* (e) Incoming DSCP/IP precedence	Class Based Weighted Fair Queuing (CBWFQ) within the MQC framework and the features below, on a per-class basis or outside of the class-based model.	Cisco Express Forwarding (CEF)	Class Based Weighted Fair Queuing (CBWFQ) within the MQC framework and the features below, on a per-class basis or outside of the class-based model.	Class-based accounting MIB and CLI
CAR and dCAR	LLQ and dLLQ		LLQ and dLLQ	
*QPPB**	WRED and dWRED		WRED and dWRED	
	CAR and dCAR		CAR and dCAR	
	FRTS and dFRTS		FRTS and dFRTS	
	GTS and dGTS		GTS and dGTS	

Looking at the basic Traffic Conditioning Mechanisms in detail:

Policing/Committed Access Rate (CAR)

The simplest concept in traffic conditioning (and in providing PHB for AF classes in the core of a DS-Domain), packets are metered, and different actions are taken, depending if the packet in question conforms, violates, or exceeds the configured average-rate, committed burst (Bc), or excess burst (Be) [Ref-E]. A packet can simply be transmitted, dropped, or remarked with a different DSCP value (moving it into a lower AF class, or changing its drop precedence value), depending on the configured policy.

Shaping (GTS/FRTS)

Sometimes, it is better to buffer packets than simply drop them in the case of congestion—especially for UDP-based applications. This can be done generically, by configuring an average-rate, committed burst (Bc), and excess burst (Be) (just as in configuring CAR). However, the biggest difference between CAR and GTS is that packets are buffered in case of congestion in shaping them. FRTS can also be employed, to have the traffic slow down when there is congestion reported by the Frame-Relay Switch.

Per Hop Behaviors (PHBs)

As the packet leaves the Ingress router, and into the network core, PHBs are enforced, depending on the packet marking with the appropriate DSCP. In Cisco IOS software, EF can be implemented using LLQ (Low Latency Queuing). AFxy PHBs can be implemented using a combination of CBWFQ (Class Based Weighted Fair Queuing) [Ref-E], and WRED (Weighted Random Early Detect) [Ref-F]/CAR.

Low Latency Queuing (LLQ) for the EF PHB

Delay sensitive traffic such as VoIP need to be given strict priority, all along the packet path. To make this happen, you can use LLQ at each hop. To ensure that excess voice traffic does not interfere with traffic of other classes, this priority queue is policed (please see section on policer above).

CBWFQ and WRED for the AF PHB

Class Based Weighted Fair Queuing using the MQC allows you to carve out bandwidth among the various classes defined. Bandwidth may be allocated to each class on an absolute basis (specified in Kbps), or as a percentage of the [sub] interface bandwidth (to which this policy will be applied). Within an AF class, packets can be dropped based on the drop precedence scheme using Weighted Random Early Detect (WRED).

Policer/CAR for the AF PHB

The policer, as described in the section above can be used to implement the PHB in the core as well. Even if packets of a class are policed at the edge of a network, the core will have many streams of a particular class merging from its numerous input interfaces, and hence will need to police the class further (at a higher aggregate rate).

Thus, in implementing DiffServ using Cisco IOS software, you define class maps and create the policy maps using the defined class maps. Finally, you apply the policy on the desired interface (or sub-interface) in either the incoming or outgoing direction.

The class maps are used to classify packets into one or more Behavior Aggregates. For example, the following classes may be defined on a DS-node:

```
                      class-map VoIP-EF
                 << Match all VoIP packets >>
                      class-map Gold-AF1
<< Match packets with DSCP value 001010 or 001100 or 001110 >>
                     class-map Silver-AF2
<< Match all traffic that is HTTP, and Citrix (using NBAR*) >>
                     class-map Bronze-AF3
             << Match all traffic that is FTP >>
                   class-map BestEffort-AF4
     << Match all other traffic, other than ones above >>
```

Note: Network Based Application Recognition (NBAR), is another powerful method in Cisco IOS software to identify traffic streams that use variable TCP/UDP ports—such as in Citrix.

In the policy map, mechanisms such as WRED (Weighted Random Early Detect), CAR (policing), GTS (Generic Traffic Shaping), FRTS (Frame-Relay Traffic Shaping), LLQ (Low Latency Queuing for traffic such as VoIP) can be specified for each class. Further, on the Cisco 7500 platforms, VIP-based distributed CAR, LLQ, GTS, WRED, and FRTS are available to offload these algorithms from the main processor, and achieve high-end scalability. These mechanisms enable Traffic Conditioning at the edge of a DS-Domain, or PHBs in a DS Internal Node. For example, the following policy may be defined on the classes defined above:

Policy-map DiffServ-Premium-and-Olympic-Policy

```
                                          class-map VOIP-EF
                              <<Strict Priority Queuing up to 128Kbps >>
                                         class-map Gold-AF1
     << Policing with excess traffic re-marked with DSCP AF13 and violate traffic dropped, and 50
                           percent of the available bandwidth allocated >>
                                        class-map Silver-AF2
     << Policing with excess traffic re-marked with DSCP AF23 and violate traffic dropped, and 25
                           percent of the available bandwidth allocated >>
                                        class-map Bronze-AF3
     << Policing with excess traffic re-marked with DSCP AF33 and violate traffic dropped, and 10
                           percent of the available bandwidth allocated >>
                                      class-map BestEffort-AF4
                   << Bandwidth available after servicing classes EF through AF3 >>
```

Note: The policer behavior above is compliant with RFC-2597. Traffic that is within the Token Bucket parameter Bc (configured burst) in an interval, is within the configured access rate, traffic between Bc and Be is excess traffic, and traffic that is more than Bc + Be (excess burst) is violate traffic that will be dropped.)

Finally, the policy can be applied on an interface or sub-interface, on an incoming or outgoing basis. For example:

```
Interface Serial1
  Service-policy output DiffServ-Premium-and-Olympic-Policy
```

In certain cases, a full-blown policy implementation is not necessary as shown above. An example might be a small sub-network, where only policing of all traffic is necessary, without any requirements for classification. For these situations, Cisco IOS software allows usage of CAR, WRED, GTS, FRTS, QPPB, and other mechanisms directly on an interface or sub-interface. However, the policy-based mechanism is a much simpler, cleaner, and scalable way of implementing DiffServ.

Cisco IOS software—Value Added Services

In addition to be providing all the core DiffServ functionality, Cisco IOS software makes it possible to define arbitrary DSCP values (local use) and associate almost any kind of policy with them. For example, “HTTP” flows between Subnet A and B may be categorized into a BA with a DSCP value of “100011” and provided 100 Kbps of bandwidth end-to-end. The IETF has divided the possible 64 DSCP values into three pools (RFC-2598) as show in the Table 3.

Table 3 The DSCP Pools

Pool	Codepoint Space	Assignment Policy
1	XXXXX0	Standards Action (EF, AFxy, Default, Class-Selector Codepoints)
2	XXXX11	Experimental/Local Usage
3	XXXX01	Experimental/Local Usage/Future Standards

Note: Any value from Pool #1, #2, or #3 can be used. Further, packets can be classified and marked using the existing IP-Precedence technique as well.

Enabling Services with Cisco IOS DiffServ

A service model applicable to typical modern networks with a combination of delay-sensitive (VoIP, real-time apps, etc.), bandwidth-sensitive (TCP, FTP, HTTP, H.323 video, etc.), and loss-sensitive (TCP, UDP, etc.) traffic is the Premium + Olympic model. The Olympic Model, as the name implies, divides traffic into Gold, Silver, and Bronze Classes with the Gold Class being more important than the Silver, than the Bronze Class. A variety of techniques (as described previously) can be used to implement this policy. Combined with best-effort service, this model can be conveniently called the Olympic+ model.

New World Opportunities

Enterprises that deploy DiffServ stand to gain tremendously by being able to deploy QoS quickly and easily in the network. Business critical and multimedia applications can be prioritized appropriately. The IP network can be transformed from a best-effort framework to a rich DiffServ region.

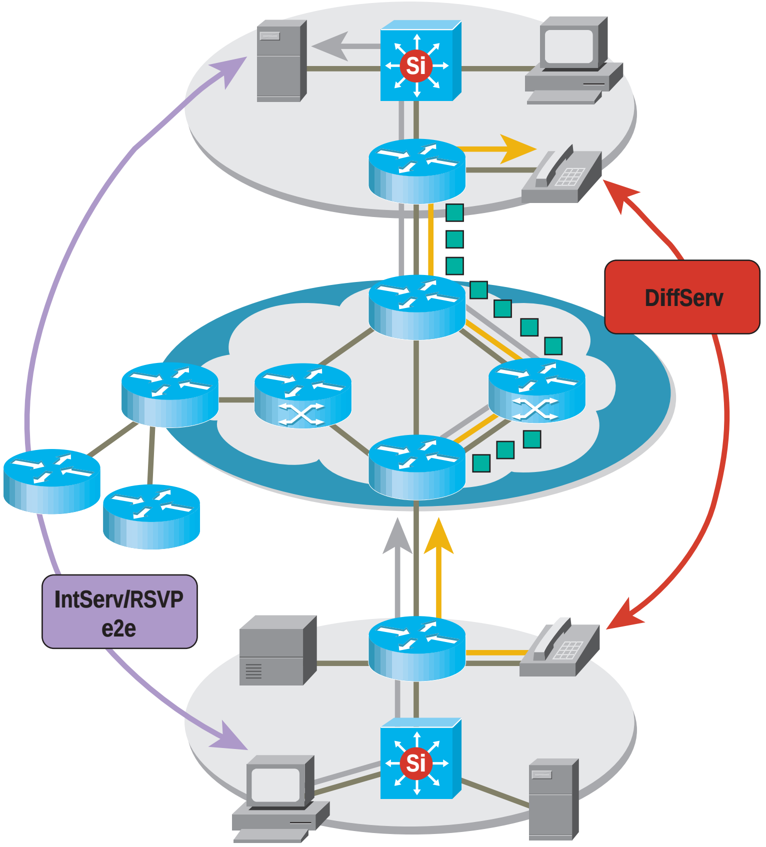
Service Providers offering a combination of QoS and VPN services stand to profit and gain the competitive edge. Cisco is committed to providing a tight integration between DiffServ and Multi Protocol Label Switching (MPLS), and enabling differentiated services over an MPLS cloud. Several MPLS-DiffServ features are already available, and more are on the way. [Ref-G].

Conclusion

Today

Putting it all together, Cisco IOS software allows IntServ and DiffServ to co-exist as two models for End-to-End QoS as shown in Figure 6. The DiffServ Domain passes the reservation requests transparently, while providing policy-based PHBs through it. The devices outside of the DS-Domain reserve bandwidth using RSVP.

Figure 6 IntServ Over DiffServ Today



DiffServ Issues—The Challenges

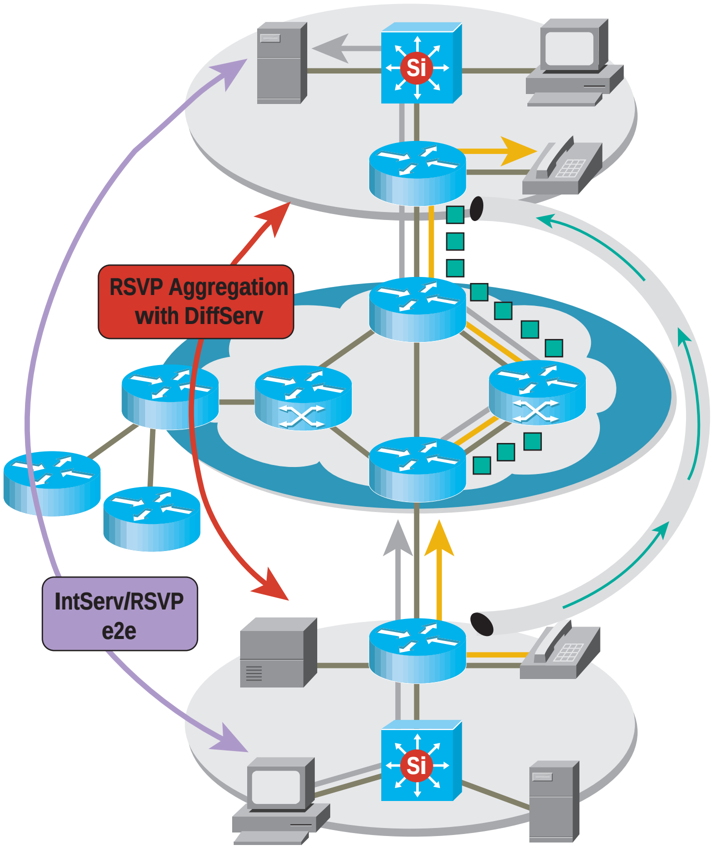
DiffServ, as great as it is, in enabling scalable and coarse-grained QoS throughout the network, has some drawbacks. In looking at them, we can see the challenges for tomorrow, and opportunities for enhancements and simplification of QoS delivery in an Internetwork:

- Provisioning—Unlike RSVP/IntServ, DiffServ needs to be provisioned. Setting up the various classes throughout the network requires knowledge of the applications, and traffic statistics for aggregates of traffic on the network. This process of application discovery and profiling can be time-consuming, although tools such as NBAR application discovery, Protocol Analyzers, and RMON (Remote Monitoring) probes can make life a bit easier.
- Billing and Monitoring—Management is still a big issue. Even though packets/sec, bytes/sec, and many other counters are available via the class-based Management Information Base (MIB), billing and monitoring are still difficult issues. For example, it may not be sufficient to prove to a customer that 9 million VoIP packets got the EF PHB treatment at all times, since it is possible that the qualitative nature of the calls that the customer made were very poor.
- Loss of Granularity—Even though QoS assurances are being made at the Class Level, it may be necessary to drill down to the flow-level to provide the requisite QoS. For example, although all HTTP traffic may have been classified as Gold, and a bandwidth of 100Mbps assigned to it, there is no inherent mechanism to ensure that a single flow does not use up that allocated bandwidth. As also, it is not easy (although not impossible) to ensure that the manufacturing department’s HTTP traffic gets priority before other department’s HTTP traffic. The MQC allows you to define hierarchical policies to accomplish this. However, it is not generic in being able to control things at a flow/granular level.
- QoS and Routing—One of the biggest drawbacks of both the IntServ and DiffServ models comes from the fact that signaling/provisioning happens separate from the routing process. Thus, there may exist a path (other than the non-default Interior Gateway Protocol [IGP], such as OSPF, ISIS, EIGRP, and so on)/Exterior Gateway Protocol (EGP, such as BGP-4, path) in the network that has the required resources, even when RSVP/DiffServ fails to find the resources. This is where traffic engineering and MPLS come into play. True QoS, with maximum network utilization will arrive with the marriage of traditional QoS and routing.

Tomorrow

In the near future, Cisco IOS software will support full RSVP aggregation, allowing reservation through a DS-Domain, and mapping of the reservation to a DSCP and PHB. The reservations will be “fat pipes” that change very slowly. This aggregated reservation overcomes the problems of maintaining thousands of RSVP soft states in the routers and flooding of refresh messages as shown in Figure 7.

Figure 7 IntServ Over DiffServ Tomorrow



RSVP Aggregation:

- For large scale deployment in the core where topology aware admission control is required inside Core
- Multiple reservations aggregated into a single Aggregate Reservation
- Aggregate Reservation is fat, slowly adjusting
- Reduced states, and Reduced signaling in core
- Aggregate Reservation mapped to a DSCP/PHB

In addition, the DiffServ challenges as discussed above will be solved, and a seamless integration achieved between DiffServ and MPLS.

Conclusion

For seamless QoS, with complete management, provisioning, and signaling support, the entire network needs to be an efficient ecosystem. Applications, hosts, switches, routers, and other network entities will all be aware of the concept of QoS, at various levels. DiffServ is the second step (following RSVP/IntServ) in making end-to-end QoS a seamless reality.

References

[Ref-A]
http://www.atmforum.com

[Ref-B]
http://www.frforum.com

[Ref-C]
http://www.ietf.org/html.charters/diffserv-charter.html

[Ref-D]
http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/120newft/120limit/120xe/120xe5/mqc/

[Ref-E]
http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/12cgcr/qos_c/qcpart4/qcpolts.htm

[Ref-F]
http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121cgcr/qos_c/qcprt3/index.htm

For additional details and information on DiffServ and Cisco IOS QoS, please refer to the following URLs:

- MQC or the Modular QoS CLI is the tool to encode QoS policy in Cisco IOS software.
  http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/120newft/120limit/120xe/120xe5/mqc/mcli.htm
- NBAR of Network based Application Recognition is a very powerful method to classify a variety of application traffic.
  http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121newft/121t/121t5/dtnbar.htm
- QPPB (QoS Policy Propagation via BGP) is another technique that can classify packets based on the community string, AS-Path, or IP access control list (ACL) in a BGP environment. Packets can be associated with different precedence/ DSCP values.
  http://www.cisco.com/univercd/cc/td/doc/product/software/ios111/cc111/bgpprop.htm
- Cisco IOS MPLS-QoS
  http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121cgcr/qos_c/index.htm
  http://www.cisco.com/go/mpls
  http://www.cisco.com/univercd/cc/td/doc/product/software/ios121/121newft/121t/121t5/mct1214t.htm

Contact Information

Additional information about Cisco IOS QoS technology can be found at http://www.cisco.com or by contacting your local Cisco representative.

DiffServ Glossary

ABR:	Available Bit Rate
AF:	Assured Forwarding
ATM:	Asynchronous Transfer Mode
BA:	Behavior Aggregate
Bc:	committed Burst
Be:	excess Burst
BGP:	Border Gateway Protocol
CAC:	Connection Admission Control
CAR:	Committed Access Rate
CBWFQ:	Class Based Weighted Fair Queuing
CIR:	Committed Information Rate
CoPS:	Common Open Policy Server
CoS:	Classification on Flows
DiffServ:	Differentiated Services
DS:	Differentiated Services
DSCP:	Differentiated Services Code Point
DTR:	data terminal ready
EF:	Expedited Forwarding
EGP:	Exterior Gateway Protocol
EIGRP:	Interior Gateway Routing Protocol
ERP:	Enterprise Resource Planning
FRF:	Frame-Relay Forum
FRTS:	Frame Relay Traffic Shaping
FRTS:	Frame-Relay Traffic Shaping
FTP:	File Transfer Protocol
GTS:	Generic Traffic Shaping
IEFT:	Internet Engineering Task Force
IGP:	Interior Gateway Protocol
IntServ:	Integrated Services
IP:	Internet Protocol
IS-IS:	Intermediate System-to-Intermediate System
ITU-T:	International Union for Telecommunications, telecommunications
LAN:	Local Area Network
LLQ:	Low Latency Queuing

MCR:	Minimum Cell Rate
MIB:	Management Information Base
MPLS:	Multi Protocol Label Switching
MQC:	Modular QoS CLI
OSPF:	Open Shortest Path First.
PCR:	Peak Cell Rate
PHB:	Per-Hop Behavior
QoS:	Quality of Service
RFCs:	Request for Comments
RMON:	Remote Monitoring
RSVP:	Resource Reservation Protocol
SLA:	Service Level Agreement
SVC:	Switched Virtual Circuit
TCP:	Transfer Control Protocol
ToS:	Type of Service
UDP:	User Datagram Protocol
VBR-rt:	Variable Bit Rate, Real-Time
VoIP:	Voice over IP
WRED:	Weighted Randomly Early Detected

Cisco Systems

Corporate Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 526-4100

European Headquarters
Cisco Systems Europe
11, Rue Camille Desmoulins
92782 Issy Les Moulineaux
Cedex 9
France
www.cisco.com
Tel: 33 1 58 04 60 00
Fax: 33 1 58 04 61 00

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-7660
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems Australia, Pty., Ltd
Level 17, 99 Walker Street
North Sydney
NSW 2059 Australia
www.cisco.com
Tel: +61 2 8448 7100
Fax: +61 2 9957 4350

Cisco Systems has more than 200 offices in the following countries. Addresses, phone numbers, and fax numbers are listed on the Cisco.com Web site at www.cisco.com/go/offices.

Argentina • Australia • Austria • Belgium • Brazil • Bulgaria • Canada • Chile • China • Colombia • Costa Rica • Croatia • Czech Republic • Denmark • Dubai, UAE Finland • France • Germany • Greece • Hong Kong • Hungary • India • Indonesia • Ireland • Israel • Italy • Japan • Korea • Luxembourg • Malaysia • Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland • Portugal • Puerto Rico • Romania • Russia • Saudi Arabia • Scotland • Singapore • Slovakia Slovenia • South Africa • Spain • Sweden • Switzerland • Taiwan • Thailand • Turkey • Ukraine • United Kingdom • United States • Venezuela • Vietnam • Zimbabwe